Intracellular gene transfer: Reduced hydrophobicity facilitates gene transfer for subunit 2 of cytochrome *c* oxidase

Daniel O. Daley*, Rachel Clifton, and James Whelan†

Plant Molecular Biology Group, Department of Biochemistry and Molecular Biology, School of Biomedical and Chemical Sciences, University of Western Australia, 35 Stirling Highway, Crawley 6009, WA, Australia

Communicated by Jeffrey D. Palmer, Indiana University, Bloomington, IN, June 12, 2002 (received for review February 13, 2002)

Subunit 2 of cytochrome *c* oxidase (Cox2) in legumes offers a rare opportunity to investigate factors necessary for successful gene transfer of a hydrophobic protein that is usually mitochondrial-encoded. We found that changes in local hydrophobicity were necessary to allow import of this nuclear-encoded protein into mitochondria. All legume species containing both a mitochondrial and nuclear encoded Cox2 displayed a similar pattern, with a large decrease in hydrophobicity evident in the first transmembrane region of the nuclear encoded protein compared with the organelle-encoded protein. Mitochondrial-encoded Cox2 could not be imported into mitochondria under the direction of the mitochondrial targeting sequence that readily supports the import of nuclear encoded Cox2. Removal of the first transmembrane region promotes import ability of the mitochondrial-encoded Cox2. Changing just two amino acids in the first transmembrane region of mitochondrial-encoded Cox2 to the corresponding amino acids in the nuclear encoded Cox2 also promotes import ability, whereas changing the same two amino acids in the nuclear encoded Cox2 to what they are in the mitochondrial-encoded copy prevents import. Therefore, changes in amino acids in the mature protein were necessary and sufficient for gene transfer to allow import under the direction of an appropriate signal to achieve the functional topology of Cox2.

The majority of mitochondrial proteins are encoded in the nuclear genome. The widely accepted endosymbiotic hypothesis proposes that genes for mitochondrial proteins were transferred from the original mitochondrial endosymbiote, and their gene products are now synthesized on cytosolic ribosomes and imported to their functional location within the mitochondrion (1). A small number of mitochondrial genes have not been transferred to the nucleus and several reasons have been proposed to account for the retention of this residual coding capacity (for reviews, see refs. 2–5). The ''hydrophobicity hypothesis'' postulates that certain mitochondrial genes encode hydrophobic proteins, which may be problematic for cellular targeting systems (6, 7). It has been suggested that hydrophobic proteins may either be mistargeted to the endoplasmic reticulum (6), and/or that long stretches of hydrophobic residues may be problematic to translocate and achieve correct orientation in the target membrane (7, 8).

The hydrophobicity hypothesis is based on observations that mitochondrial proteins encoded in the nuclear genome contain fewer hydrophobic stretches, that these stretches have a lower hydrophobicity score, and that they are positioned more proximal to the C-terminal domain of the protein (6–8). Apocytochrome *b* (Cob) and cytochrome *c* oxidase subunit 1 (Cox1) contain 8 and 12 transmembrane regions (TM), respectively (8, 9), and are the only genes universally encoded in mitochondrial genomes (3). Furthermore, it has been experimentally demonstrated that Cob cannot be imported into the mitochondrion when attached to a mitochondrial targeting presequence (8).

In plants and green algae, the transfer of genes from the mitochondrial to the nuclear genome is an ongoing process as evidenced from a number of evolutionarily recent gene transfer events (for review, see ref. 2). Some of these events have involved hydrophobic inner membrane proteins (10–17), suggesting that the proposed hydrophobicity problem can be overcome. In the cases where the gene for a hydrophobic protein has been transferred to the nucleus, it was noted that the product of the nuclear gene was less hydrophobic than its mitochondrial counterpart. However, there is no experimental evidence to determine whether such changes in hydrophobicity were necessary to facilitate the gene transfer.

The recent gene transfer events characterized in plants and green algae offer a unique opportunity to determine whether a reduction in hydrophobicity of transmembrane segments was necessary for successful gene transfer. Any changes that occurred after gene transfer are evident, and the effect of such changes can be tested experimentally. Such a situation exists for *cox2* in legumes, where active nuclear and mitochondrial-encoded copies exist in some species (15–18). This situation is open to experimental manipulation to determine whether changes in hydrophobicity that have occurred after gene transfer to the nucleus played a role in facilitating import of the protein into the mitochondrion, and thus successful gene transfer. In this study, we have analyzed the sequences for nuclear and mitochondrial-encoded Cox2 in soybean (*Glycine max*), and other legumes, and determined that a decrease in hydrophobicity in the first transmembrane spanning region was necessary for mitochondrial import of the nuclear encoded protein.

Materials and Methods

DNA Manipulations. The nuclear encoded *Cox2* cDNA clone from *Glycine max* (n*GmCox2*) was obtained as described (16). The mitochondrial-encoded *cox2* cDNA clone from *Amphicarpa bracteata* (mt*Abcox2*) was obtained from K. A. Adams and J. D. Palmer. The mitochondrial-encoded *cox2* cDNA clone from *Glycine max* (mt*Gmcox2*) was obtained by site-directed mutagenesis of 13 different residues in mt*Abcox2* (Fig. 1*C*), so that it resembled the predicted edited version of the mitochondrial *Glycine max* gene. Cytidine to uridine editing sites for mt*Gmcox2* were chosen based on conserved Cytidine to Uridine editing sites for pea and cowpea *cox2* (17, 18). Site-directed mutagenesis was performed by using the QuikChange Site-Directed Mutagenesis kit (Stratagene, Sydney), according to the manufacturer's instructions. All constructs and mutants were verified by DNA sequencing using an ABI 310 genetic analyzer (Perkin–Elmer, Melbourne).

Abbreviations: Cox1, subunit 1 of cytochrome *c* oxidase; Cox2, subunit 2 of cytochrome *c* oxidase; Cob, apocytochrome *b*; KD, Kyte and Doolittle; TM, transmembrane region; aWW, augmented Whitely White; MSS, mitochondrial sorting sequence.

*Present address: Department of Biochemistry and Biophysics, Stockholm University, S-10691 Stockholm, Sweden.

†To whom reprint requests should be addressed. E-mail: seamus@cyllene.uwa.edu.au.

 www.pnas.org/cgi/doi/10.1073/pnas.122354399

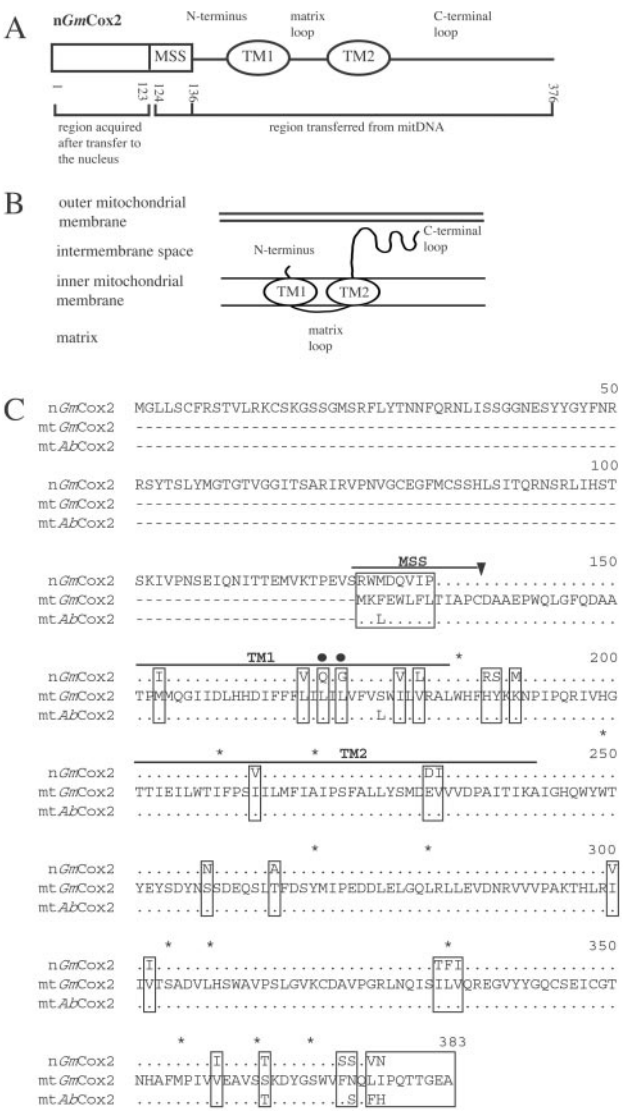

Fig. 1. Cytochrome *c* oxidase subunit 2 from soybean. (*A*) Domain-structure of the nuclear-encoded Cox2 from soybean, showing the region that was transferred from the mitochondrial genome and the targeting presequence that was acquired on integration into the nuclear genome. MSS is for mitochondrial sorting signal, and TM is for transmembrane spanning domains. (*B*) Topology of Cox2 in the inner mitochondrial membrane based on homology with yeast and bovine where the crystal structure has been determined. (*C*) Amino acid alignment of the nuclear-encoded Cox2 (n*Gm*Cox2) from soybean and mitochondrial-encoded Cox2 from soybean (mt*Gm*Cox2) and *Amphicarpa bracteata* (mt*Ab*Cox2). Amino acid differences between the nuclear and organelle-encoded soybean proteins are boxed and TM regions are indicated. C to U editing sites in the mt*Gm*Cox2 sequence are indicated by an asterisk. Residues 169 and 171, which were changed are indicated with a filled circle (●). The start of the mature Cox2 protein, as previously experimentally determined, is indicated with a vertical arrow.

Hydrophobicity Analysis. Hydrophobicity analysis of Cox2 sequences was performed by using the scales of Kyte and Doolittle (KD) (19) and Jayasinghe *et al.* (20) using default parameters. Local hydrophobicity ($\langle H \rangle$) was calculated by using a scanning window of 17 and 19 residues, respectively. The average regional hydrophobicity ($\langle H \rangle_{60-80}$), termed mesohydrophobicity, was calculated by scanning for the average maximum hydrophobicity using scanning windows from 60–80 residues and averaging the values, as described by Claros *et al.* (8) using the MITOPROT V1.0 program (21).

Mitochondrial Isolation, *in Vitro* Import Assays, and Western Blotting. Mitochondria were prepared from 5-day-old soybean cotyledons as described by Day *et al.* (22). [^{35}S]-labeled precursor proteins were synthesized from cDNA clones, and *in vitro* import assays were carried out as previously described and modified by Daley *at al.* (16, 23). Osmotic swelling to rupture the outer mitochondrial membrane of mitochondria was performed as described by Weinhues *et al.* (24). Proteins were separated by 12% SDS/PAGE, gels were dried and exposed to a BAS TR2040S plate for 24 h. Detection was carried out on a BAS 2500 according to the manufacturers' instructions (Fuji, Tokyo). Import efficiencies were calculated as the amount of protease protected product relative to the amount of added precursor protein, and expressed relative to the wild-type n*Gm*Cox2, which was set to 100%. Western blotting was performed as described by Daley *et al.* (16), using antisera raised to cytochrome *c* (PharMingen) the mitochondrial uncoupling protein (25), Hsp70 (E. Galser, Stockholm Univ., Stockholm) and Porin (Tom Elthon, Univ. of Nebraska, Lincoln).

Results

Amino Acid Differences Between Nuclear and Mitochondrial-Encoded Cox2 Change the Hydropathicity Profile of the Protein. The import of nuclear-encoded Cox2 from soybean (n*Gm*Cox2[‡]) has been shown to require a unique mitochondrial targeting presequence of 136 amino acid residues for mitochondrial import. This targeting presequence consists of 13 amino acids that were present from the mitochondrial-encoded Cox2 (mt*Gm*Cox2), and 123 amino acids that were acquired on integration into the nuclear genome (Fig. 1*A*) (16). This extremely long presequence is processed three times in the import route taken by n*Gm*Cox2, and all regions are necessary for correct maturation. Thus, it appears that the mature region of n*Gm*Cox2 contains no targeting information (16). Therefore the $N_{out}-C_{out}-$ topology in the inner membrane (Fig. 1*B*) (9, 26) is most likely achieved by translocation of one or both TM across the inner membrane and reinsertion back into the membrane from the matrix side during mitochondrial import.

Comparison of mt*Gm*Cox2 with n*Gm*Cox2 reveals a number of amino acid changes in the mature protein that occurred after gene transfer to the nucleus (Fig. 1*C*). Plant mitochondrial-encoded genes undergo extensive post-transcriptional C to U editing, at sites that are well conserved between species. As the *cox2* transcript is not edited in soybean mitochondria, it is a pseudogene (15). For the purposes of this study, we changed the soybean *cox2* gene at C to U editing sites by using site-directed mutagenesis based on the conserved editing patterns from pea and cowpea (17, 18, 27) so that the gene product should reflect an authentic Cox2. Twenty-five amino acid differences were evident between the two predicted mature proteins (Fig. 1*C*), and were distributed throughout the coding region of n*Gm*Cox2. In only one instance did an amino acid difference coincide with an editing site, strongly suggesting that the changes in amino acids observed between the organellar and nuclear encoded copies were not caused by transfer of an unedited or partially edited transcript. There were no changes in those residues that have been shown to be involved in copper binding, electron transport, and cytochrome *c* docking in other species (data not shown). In several legume species that contain dual expressed Cox2, a similar pattern of amino acid differences was observed between the nuclear and mitochondrial-encoded proteins. This

[‡]The distinction between nuclear- and mitochondrial-encoded proteins will be indicated by an n (nuclear-encoded) and mt (mitochondrial-encoded). The organism will be indicated by the first letters of the species name (i.e., *Glycine max* = *Gm*). Chimeric constructs will be designated with an n or mt after each region to indicate if taken from a nuclear or mitochondrial gene, respectively.

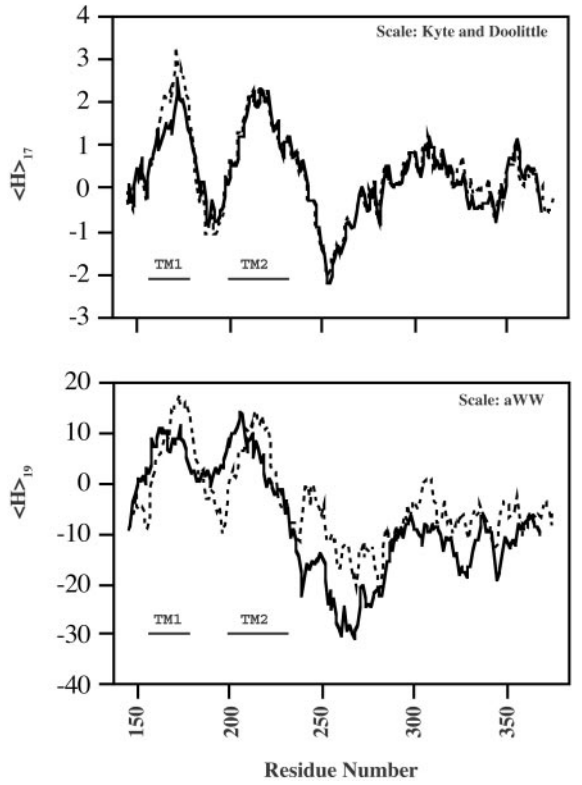

Fig. 2. Comparison of hydropathicity profiles for the mitochondrial-encoded and nuclear encoded subunit 2 of cytochrome *c* oxidase from soybean. Hydropathicity was calculated by using a sliding window of 17 or 19 residues (default values) respectively for KD (19) and aWW (20). The position of the two transmembrane regions, TM1 and TM2, are indicated with solid lines. mt*Gm*Cox2 is indicated with a dashed line, and n*Gm*Cox2 is indicated with a solid line. Only a solid line is visible where the lines overlap, indicating the hydropathy plots overlap.

observation supported our choice of editing pattern used on mt*Gmcox2*.

We investigated whether the amino acid differences between mt*Gm*Cox2 and n*Gm*Cox2 altered the hydropathicity profile of the protein. When the KD scale (19) was used to determine the hydropathicity profiles, it was apparent that there was a significant reduction in local hydrophobicity ($\langle H \rangle_{17}$) for the first transmembrane spanning domain (TM1) after gene transfer to the nucleus (Fig. 2). The change in $\langle H \rangle_{17}$ for TM1 was in stark contrast to the remainder of the two proteins, which had almost identical hydropathicity profiles. In addition to the widely used KD, scale we used a newer experimentally determined aWW scale, which indicated that the reduction in hydrophobicity of TM1 for soybean was greater than 30% and conversely that there was an 80% increase in the hydrophobicity for TM2 (Fig. 2). The aWW scale takes into account the peptide bonds in the amino acid backbone and salt bridges between amino acids, in addition to the side chain characteristics used in the KD and other commonly used scales (20). Therefore this scale is likely to give a more accurate picture of the changes in hydrophobicity that have occurred. Examination of the hydrophobicity of Cox2 in other legumes where dual-expressed genes for Cox2 exist yielded a clear pattern whereby a decrease in hydrophobicity of TM1 and increase of TM2 were evident (Table 1).

Mitochondrial Import of Cox2 in Legumes Required a Reduction in Hydrophobicity After Gene Transfer to the Nucleus. Two fusion proteins were constructed to determine whether the amino acid differences, and corresponding change in hydropathicity profile, between mt*Gm*Cox2 and n*Gm*Cox2 were necessary to facilitate mitochondrial import. n*Gm*1-123/mt*Gm*124-383 (Fig. 3*A*) contained the mitochondrial targeting presequence of n*Gm*Cox2 attached to mt*Gm*Cox2 (the region that was originally transferred from the mitochondrial genome). Additionally, the mitochondrial-encoded Cox2 protein from *Amphicarpa bracteata* (mt*Ab*Cox2) was attached to the acquired presequence of n*Gm*Cox2 to give the n*Gm*1-123/mt*Ab*124-376 construct (Fig. 3*A*). As mt*Ab*Cox2 represents a functional protein in a species that is closely related to soybean (15), it provides a control that is not dependent on predicting C to U editing patterns.

Table 1. Local hydrophobicity ($\langle H \rangle$) values for Cox2 from those legumes with both mitochondrial- and nuclear-encoded copies

		$\langle H \rangle_{17}$ KD		$\langle H \rangle_{19}$ aWW	
Species	Gene	TM1	TM2	TM1	TM2
Soybean (*Gm*)	mt	3.253	2.306	5.83	0.99
	n	2.559	2.288	3.82	1.78
Amphicarpea bracteata (*Ab*)	mt	3.524	2.306	6.52	0.98
	n	2.024	2.265	3.03	2.15
Dumasia villosa	mt	3.524	2.306	6.52	0.98
	n	2.729	2.306	2.96	1.77
Lespedeza formosa	mt	3.524	2.306	6.52	0.98
	n	1.888	2.265	1.87	0.90
Neonotonia wighdi	mt	3.465	2.306	7.09	0.98
	n	2.712	2.306	2.92	1.49
Pseudeminia comosa	mt	3.524	2.306	6.52	0.98
	n	2.753	2.247	2.34	1.31

Values were calculated for both transmembrane spanning domains (TM1 and TM2) using a sliding window of 17 and 19 residues (default values) respectively, for Kyte and Doolittle (19) and aWW (20).

In vitro import assays were carried out to determine whether the mitochondrial-encoded Cox2 proteins could be imported and cleaved to the mature form as previously determined for n*Gm*Cox2 (16). *In vitro* import of [^{35}S]methionine-labeled proteins into mitochondria was assessed by (*i*) generation of an additional processed band in the presence of mitochondria, (*ii*) protection from externally added protease, and (*iii*) dependence on the presence of a membrane potential across the inner mitochondrial membrane. Incubation of [^{35}S]methionine-labeled n*Gm*Cox2 with isolated soybean cotyledon mitochondria followed by treatment of the mitochondria with proteinase K yielded a protein band with an apparent molecular mass of 31 kDa (Fig. 3*B*, lane 3, denoted M). We have previously shown that this corresponds to mature Cox2 protein (16). The imported product was only generated in the presence of a $\Delta\Psi$, as the addition of valinomycin to the assay inhibited protein import (Fig. 3*B*, lane 5). When [^{35}S]methionine-labeled n*Gm*1-123/mt*Gm*124-383 and n*Gm*1-123/mt*Ab*124-376 (organelle-encoded forms of Cox2) were incubated with isolated soybean mitochondria, no protease protected product was generated (Fig. 3*B*, lanes 8 and 13, respectively). It was concluded that these constructs could not be imported into the mitochondrion. The additional lower molecular mass products present in the translation mixture alone are most likely caused by initiation of translation at internal methionines, a common occurrence with *in vitro* translation lysates, and play no role in the import reaction (16).

To determine which region of the mitochondrial-encoded Cox2 protein was inhibiting mitochondrial import the transmembrane spanning domains of n*Gm*1-123/mt*Ab*124-376 were removed individually and in tandem. We assessed the intramitochondrial location of any imported proteins by rupturing the

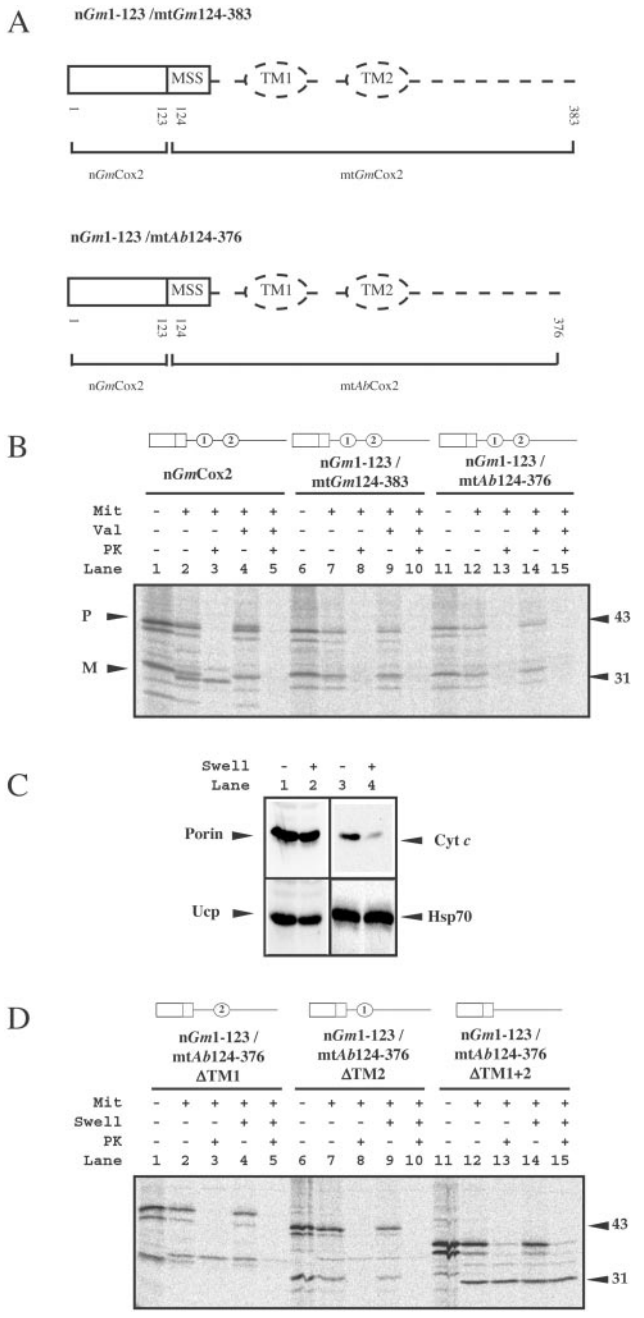

Fig. 3. The mitochondrial-encoded subunit 2 of cytochrome *c* oxidase (mt*Gm*Cox2) cannot be imported into mitochondria. (*A*) Chimeric constructs synthesized to determine whether a mitochondrial-encoded Cox2 could be imported into mitochondria. Sequence derived from the nuclear gene is marked by n, sequence derived from the mitochondrial gene is marked by mt. *Gm* indicates the sequence was obtained from soybean (*Glycine max*), and *Ab* indicates that the sequence was obtained from *Amphicarpa bracteata*. Other abbreviations are as for Fig. 1. (*B*) *In vitro* import of n*Gm*Cox2, n*Gm*1-123/mt*Gm*124-383 and n*Gm*1-123/mt*Ab*124-376 constructs into soybean cotyledon mitochondria. Lane 1, translation lysate containing n*Gm*Cox2 precursor protein (denoted P). Lane 2, precursor protein incubated with isolated soybean mitochondria. Lane 3, as lane 2 with proteinase K added. Lanes 4 and 5 are as for lanes 2 and 3, except in the presence of valinomycin to dissipate the membrane potential. Lanes 6–10 are as for lanes 1–5, except with the n*Gm*1-123/mt*Gm*124-383 precursor protein. Lanes 11–15 are as for lanes 1–5, except with the n*Gm*1-123/mt*Ab*124-376 precursor protein. Molecular weight standards are indicated to the right of the panel. The mature Cox2 is denoted M. (*C*) Western blot analysis of mitochondria and mitoplasts used in *D*, with antisera raised to porin (outer mitochondrial membrane), cyt *c* (intermembrane space), Ucp (inner mitochondrial membrane), and Hsp70 (matrix). (*D*) *In vitro* import of n*Gm*1-123/mt*Ab*124-376 without transmembrane spanning domains. Lane 1, translation lysate containing n*Gm*1-123/mt*Ab*124-376 ΔTM1 precursor protein. Lane 2, precursor protein incubated with isolated soybean cotyledon mitochondria. Lane 3, as lane 2 with proteinase K added. Lanes 4 and 5 are as for lanes 2 and 3, except that the outer mitochondrial membrane had been stripped by osmotic swelling. Lanes 6–10 are as for lanes 1–5, except with the n*Gm*1-123/mt*Ab*124-376 ΔTM2 precursor protein. Lanes 11–15 are as for lanes 1–5, except with the n*Gm*1-123/mt*Ab*124-376 ΔTM1 + 2 precursor protein. Quantitation of import efficiencies are included in Table 2. Molecular weight standards are indicated to the right of the panel.

outer mitochondrial membrane by osmotic swelling, allowing the externally added protease access to the intermembrane space but not the mitochondrial matrix (Fig. 3*C*). Removal of TM1 alone (n*Gm*1-123/mt*Ab*124-376 ΔTM1) allowed the import of mt*Ab*Cox2 (Fig. 3*D*, lane 3), with an increased efficiency relative to n*Gm*Cox2 (Table 2, which is published as supporting information on the PNAS web site, www.pnas.org). When the outer mitochondrial membrane was ruptured after the *in vitro* import assay, the imported protein was accessible to the added protease (Fig. 3*D*, lane 5) indicating that TM2 was arrested in the inner mitochondrial membrane. By contrast, in the absence of TM2 (n*Gm*1-123/mt*Ab*124-376 ΔTM2), mt*Ab*Cox2 was only imported weakly into isolated mitochondria relative to n*Gm*Cox2 (Fig. 3*D*, lane 8). The weakly imported product was arrested in the inner membrane as evidenced by protease digestion when the outer membrane was ruptured, this time because of the presence of the TM1 domain (Fig. 3*D*, lane 10). Deletion of both TM1 and TM2 (n*Gm*1-123/mt*Ab*124-376 ΔTM1 + 2) allowed efficient import of the organelle-encoded Cox2 into the matrix as the protein was protected when the outer membrane was ruptured (Fig. 3*D*, lanes 13 and 15). Taken together these data indicate that organelle-encoded TM1 is not only inhibitory to import but that it will not pass through the inner membrane, thus preventing Cox2 from reaching the correct topology necessary for assembly and function.

A Reduction in Hydrophobicity of TM1 Was Necessary to Enable Mitochondrial Import of nGmCox2. We undertook site-directed mutagenesis of the TM1 domain n*Gm*1-123/mt*Gm*124-383 in an attempt to reconstruct the molecular events, after gene transfer, that lead to mitochondrial import. Of the 6 amino acid differences between the TM1 domains of n*Gm*Cox2 and mt*Gm*Cox2 (Fig. 1*C*), only 2 contributed significantly to the overall decrease in ⟨H⟩ for this region. We changed only these amino acids in the n*Gm*1-123/mt*Gm*124-383 construct to the corresponding residue in n*Gm*Cox2 (Table 3, which is published as supporting information on the PNAS web site). By simulating the amino acid change after transfer at residue 169 [i.e., from leucine (L) to glutamine (Q)], there was a decrease in the ⟨H⟩ of TM1 that was reflected in both the KD (19) and aWW (20) hydrophobicity scales (Table 3). However, the amino acid change at residue 169 alone would not permit mitochondrial import of n*Gm*1-123/mt*Gm*124-383 (Fig. 4*A*, lane 3 vs. lane 6). We also simulated the amino acid change that occurred at residue 171 [i.e., from leucine (L) to glycine (G)]. The double amino acid change caused a further reduction in the ⟨H⟩ for TM1, to a level that was close to n*Gm*Cox2, and permitted mitochondrial import of n*Gm*1-123/mt*Gm*124-383 (Fig. 4*A*, lane 3 vs. lane 9; Table 3). Although the import efficiency of n*Gm*1-123/m*Gm*124-383$L_{169}Q/L_{171}G$ was somewhat lower than n*Gm*Cox2, it was imported and processed to the correct apparent molecular mass as n*Gm*Cox2 (16).

To verify the observation that amino acid changes at residues 169 and 171 were instrumental in enabling import of n*Gm*Cox2, we also changed these residues in n*Gm*Cox2 to their corresponding residue in mt*Gm*Cox2 to determine whether the increase in ⟨H⟩ for TM1 could inhibit mitochondrial import. Both the single

EVOLUTION

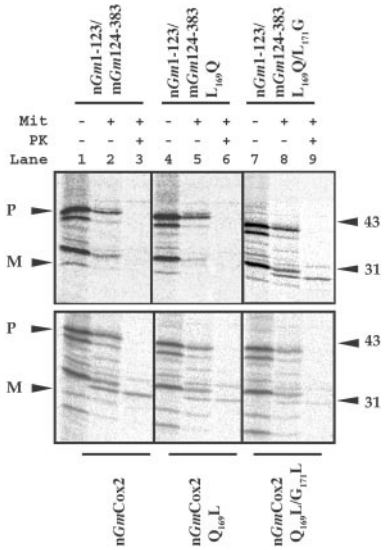

Fig. 4. Two residues in the first transmembrane spanning region of cytochrome *c* oxidase subunit 2 determine mitochondrial import. Lane 1, translation lysate containing precursor protein (denoted P). Lane 2, precursor protein incubated with isolated soybean mitochondria. Lane 3, as lane 2 with proteinase K added. Lanes 4–6 and 7–9, as for lanes 1–3 except with different precursor proteins. Precursor protein used in each instance is indicated above or below the panel. Molecular weight standards are indicated to the right of the panel.

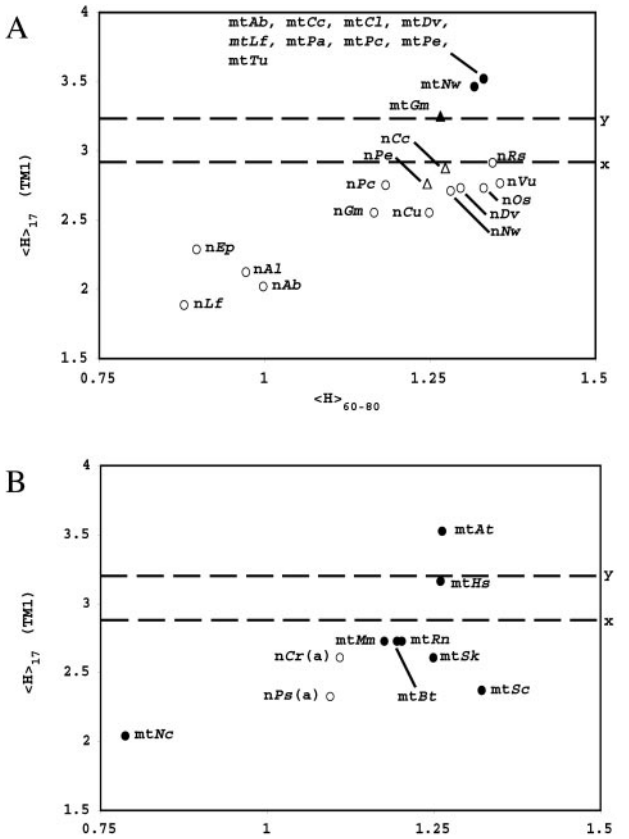

Fig. 5. Mesohydrophobicity vs. local hydrophobicity ''scatter plots'' for cytochrome *c* oxidase subunit 2 from different organisms. (*A*) Scatter plot of mitochondrial-encoded (closed symbols) and nuclear-encoded (open symbols) Cox2 sequences from legumes. Proteins are distributed on the horizontal axis according to their maximum mesohydrophobicity ($\langle\langle H\rangle_{60-80}$) and on the vertical axis according to the maximum value of local hydrophobicity ($\langle H\rangle_{17}$) for the first transmembrane spanning domain (TM1). Hydrophobicity was calculated by using the scale of KD (19). Values for functionally active proteins are indicated by circles, and protein sequences derived from pseudogenes are indicated by triangles. The maximum $\langle H\rangle_{17}$ for TM1 in legumes that is encoded by a nuclear gene, and can therefore be imported, is shown by line X. The level of $\langle H\rangle_{17}$ for TM1 in legumes that cannot be imported is show by line Y. *Ab*, *Amphicarpa bracteata*; *Al*, *Atylosia lineata*; *Cc*, *Calopogonium caeruleum*; *Cl*, *Cologonia lemmonii*; *Cu*, *Cullen*; *Dv*, *Dumasia villosa*; *Ep*, *Eriosema psoraloides*; *Gm*, *Glycine max* (soybean); *Lf*, *Lespedeza formosa*; *Nw*, *Neonotonia weightii*; *Os*, *Otholobium sericeum*; *Pa*, *Pseudovigna argentea*; *Pc*, *Pseudiminia comosa*; *Pe*, *Pachyrhizus erosus*; *Tu*, *Teramnus uncinatus*; *Vu*, *Vigna unguiculata*. (*B*) Scatter plot of mitochondrial-encoded (closed symbols) and nuclear-encoded (open symbols) Cox2 sequences from many species. Annotation as for *A*. *At*, *Arabidopsis thaliana*; *Bt*, *Bos taurus*; *Cr*, *Chlamydomonas reinhardtii*; *Hs*, *Homo sapiens*; *Mm*, *Mus muscularis*; *Nc*, *Neurospora crassa*; *Ps*, *Polymotella sp*; *Rn*, *Rattus novegicus*; *Sc*, *Saccharomyces cerevisiae*; *Sk*, *Saccharomyces kluyveri*.

and double changes at positions 169 and 171 decreased import dramatically, supporting the observations that amino acid changes in TM1 of the transferred protein were necessary to allow import (Fig. 4 and Table 3).

As 8 of 13 residues from the mitochondrial sorting sequence (MSS) of mt*Gm*Cox2 changed after gene transfer to the nucleus (Fig. 1*C*), we investigated whether these changes had any effect on the import of n*Gm*Cox2. *In vitro* import of n*Gm*Cox2 with either the nuclear or mitochondrial MSS region had little effect on the import profile of n*Gm*Cox2 (data not shown). This observation is in agreement with the previous experiments where n*Gm*1-123/mt*Ab*124-376 with either TM1, or both TM regions removed could be imported efficiently and both contained the mitochondrial MSS region (Fig. 3*D*). Thus the mitochondrial MSS region is not inhibitory to import.

Discussion

Gene transfer of *cox2* from the mitochondrion to the nucleus has only been reported twice, once in Chlamydomonad algae and once in a group of legumes (12, 15). Because *cox2* is present in the mitochondrion of most completely sequenced mitochondrial genomes (3), transfer of the *cox2* gene to the nucleus appears to be rare. In both reported transfer events the mitochondrial targeting presequence acquired was extremely long, encoding greater than 130 amino acids. We have experimentally shown that this long presequence cannot be replaced by another mitochondrial targeting signal in soybean (16). Thus, in addition to the usual requirements for gene transfer (31), Cox2 in legumes requires a unique tripartite mitochondrial targeting signal that is cleaved in a process unique to legumes (16). In this study, we show that *cox2* gene transfer in legumes requires an additional step, involving a reduction in local hydrophobicity of the first transmembrane region. Notably, a reduction in the overall hydrophobicity of the Cox2 protein, termed mesohydrophobicity, does not seem to be required (Fig. 5*A*). Although a modest decrease in mesohydrophobicity is evident between the mt*Gm*Cox2 and n*Gm*Cox2, this is not a general pattern that is evident with all legumes that contain a nuclear encoded Cox2 protein (Fig. 5*A*).

We used mitochondrial and nuclear encoded forms of the Cox2 protein from soybean, and were able to experimentally determine what specific change in the hydropathicity profile was required for import of nuclear encoded Cox2. The presence of a gene in both the mitochondrion and nucleus is rare, and thus the *cox2* gene transfer in legumes offers a unique window into this process. Analysis of hydrophobicity indicated that TM1 was a potential barrier for mitochondrial import, because of its high local hydrophobicity and the fact it had decreased in all cases where there was a nuclear encoded Cox2 protein in legumes (Table 1). We experimentally verified this observation by re-

moving TM1, which allowed import of a mitochondrial-encoded Cox2 protein. We located the changes to two specific residues by using a bidirectional approach that allows the other changes that are evident to be taken into account; i.e., with mt*Gm*Cox2 changing two residues were sufficient to promote import even though 23 other residues differ in n*Gm*Cox2, whereas in n*Gm*Cox2 changing 2 residues inhibited import. It has been reported that, after transfer, either the nuclear or mitochondrial-encoded Cox2 can become inactivated (15). Thus, the nuclear copy must rapidly become functional to avoid inactivation by the accumulation of deleterious mutations. Rapid activation requiring changing 25 amino acids as is seen between the organelle and nuclear encoded copy of soybean Cox2 would be difficult to envisage, but rapid changes in just two amino acids to allow activation would be easier. The additional differences present between the nuclear and mitochondrial-encoded copies could easily have taken place subsequent to gene activation and may increase the import efficiency of n*Gm*Cox2.

In fungal and animal mitochondria, the use of a nonuniversal genetic code may simply lock some *cox2* genes in the mitochondrion, as most organelle-encoded Cox2 proteins are below the threshold level of hydrophobicity for import (Fig. 5*B*). The yeast *Saccharomyces cerevisiae* offers an interesting example as in terms of mesohydrophobicity, yeast Cox2 does not represent an extreme example (Fig. 5*B*), and local hydrophobicity of TM1 (2.371) using the KD scale is lower than n*Gm*Cox2 (Table 1). Therefore, this protein could be potentially transferred if not for the barrier of the genetic code. Attempts to reengineer yeast *cox2* as a nuclear encoded protein do not appear to have been successful, as only a construct containing TM1 could be imported into the mitochondrial matrix (32). These attempts used a mitochondrial targeting sequence from *Neurospora crassa* subunit 9 of ATP synthase, and thus the inability of the complete protein to be imported may be caused by the overall hydrophobicity when combined with this targeting signal. Previous studies have shown that a standard mitochondrial targeting signal cannot support import of n*Gm*Cox2, similar to the findings that other mitochondrial targeting signals could not support the import of yeast Cox2 or Cob (8, 16, 27). Thus, in addition to a reduction in local hydrophobicity, which was necessary for successful gene transfer of Cox2, other barriers were also overcome in the Cox2 gene transfer in legumes.

In conclusion, changes in amino acids, which reduce the local hydrophobicity of the first transmembrane segment of Cox2, were necessary and sufficient to facilitate successful gene transfer in addition to the other unique features required for this rare gene transfer event. The ability for Cox2 to circumvent the restraints placed on it by cellular targeting systems has interesting implications for the ongoing process of organelle to nucleus gene transfer. It suggests that even those mitochondrial genes that encode for hydrophobic proteins may be ''unlocked' from the mitochondrial genome. This observation is in keeping with the hypothesis that there is a hierarchical order of gene loss for mitochondrial genes, and that those genes that encode hydrophobic proteins are simply the last to go. As *cox1* and *cob* encode more hydrophobic proteins than *cox2*, it is not surprising that they have not yet been discovered in the nucleus. It is evident from mitochondrial sequencing projects that many genes retained by mitochondrial genomes, most notably ribosomal protein genes, cannot be accounted for by the hydrophobicity hypothesis. Some genes for ribosomal proteins are still encoded by mitochondrial genomes (3) or have been transferred late in mitochondrial evolution in angiosperms (10, 33–39). For instance, *rps10* encodes a hydrophilic protein that can be imported in the absence of a mitochondrial targeting signal and poses no problems for mitochondrial import (39). It must therefore be considered that factors other than hydrophobicity preclude the transfer of a gene to the nucleus.

We thank Keith Adams and Jeff Palmer for the gift of the *Amphicarpa* cDNA clone and stimulating discussions, and Monika Murcha and Charlotta Rudhe for technical assistance. This work was supported by an Australian Research Council grant (to J.W.).

1. Gray, M. W. (1992) *Int. Rev. Cytol.* **141,** 233–357.
2. Palmer, J. D., Adams, K. L., Cho, Y., Parkinson, C. L., Qui, Y. & Song, K. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 6960–6966.
3. Lang, B. F., Gray, M. W. & Burger, G. (1999) *Annu. Rev. Genet.* **33,** 351–397.
4. Martin, W. & Herrmann, R. G. (1998) *Plant Physiol.* **118,** 9–17.
5. Doolittle, W. F. (1998) *Trends Genet.* **14,** 307–311.
6. von Heijne, G. (1987) *FEBS Lett.* **198,** 1–4.
7. Popot, J.-L. & de Vitry, C. (1990) *Annu. Rev. Biophys. Biophys. Chem.* **19,** 369–403.
8. Claros, M. G., Perea, J., Shu, Y., Samatey, F. A., Popot, J.-L. & Jacq, C. (1995) *Eur. J. Biochem.* **228,** 762–771.
9. Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Itoh-Shinzawa, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1996) *Science* **272,** 1136–1144.
10. Adams, K. L., Rosenblueth, M., Qui, Y.-L. & Palmer, J. D. (2001) *Genetics* **158,** 1289–1300.
11. Perez-Martinez, X., Vazquez-Acevedo, M., Tolkunova, E., Funes, S., Claros, M. G., Davidson, E., King, M. P. & Gonzalez-Halphen, D. (2000) *J. Biol. Chem.* **275,** 30144–30152.
12. Perez-Martinez, X., Antaramian, A., Vazquez-Acevedo, M., Funes, S., Tolkunova, E., d'Alayer, J., Claros, M. G., Davidson, E., King, M. P. & Gonzalez-Halphen, D. (2001) *J. Biol. Chem.* **276,** 11302–11309.
13. Funes, S., Davidson, E., Claros, M. G., van Lis, R., Perez-Martinez, X., Vazquez-Acevedo, M., King, M. P. & Gonzalez-Halphen, D. (2001) *J. Biol. Chem.* **277,** 6051–6058.
14. Kobayashi, Y., Knoop, V., Fukuzawa, H., Brennicke, A. & Ohyama, K. (1997) *Mol. Gen. Genet.* **256,** 589–592.
15. Adams, K. L., Song, K., Roessler, P. G., Nugent, J. M., Doyle, J. L., Doyle, J. J. & Palmer, J. D. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 13863–13868.
16. Daley, D. O., Adams, K. L., Clifton, R., Qualmann, S., Millar, A. H., Palmer, J. D., Pratje, E. & Whelan, J. (2002) *Plant J.* **30,** 11–21.
17. Covello, P. S. & Gray, M. W. (1992) *EMBO J.* **11,** 3815–3820.
18. Nugent, J. M. & Palmer, J. D. (1991) *Cell* **66,** 473–481.
19. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
20. Jayasinghe, S., Hristova, K. & White, S. H. (2001) *J. Mol. Biol.* **312,** 927–934.
21. Claros, M. G. (1995) *Comput. Appl. Biosci.* **11,** 441–447.
22. Day, D. A., Neuberger, M. & Douce, R. (1985) *Aust. J. Plant Physiol.* **12,** 219–228.
23. Whelan, J., Hugosson, M., Glaser, E. & Day, D. A. (1995) *Plant Mol. Biol.* **27,** 769–778.
24. Weinhues, U., Koll, H., Becker, K., Guiard, B. & Hartl, F.-U. (1992) in *Protein Targeting: A Practical Approach*, eds. Magee, A. I. & Wileman, T. (Oxford Univ. Press, Oxford), pp. 135–159.
25. Considine, M. J., Daley, D. O. & Whelan, J. (2001) *Plant Physiol.* **126,** 1619–1629.
26. Bisson, R., Steffens, G. C. M. & Buse, G. (1982) *J. Biol. Chem.* **257,** 6716–6720.
27. Grabau, E. A. (1987) *Curr. Genet.* **11,** 287–293.
28. Andersson, S. G. E. & Kurland, C. G. (1998) *Trends Microbiol.* **6,** 263–268.
29. Berg, O. G. & Kurland, C. G. (2000) *Mol. Biol. Evol.* **17,** 951–961.
30. Blanchard, J. L. & Lynch, M. (2000) *Trends Genet.* **16,** 315–320.
31. Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. & Schuster, W. (1993) *FEBS Lett.* **325,** 140–145.
32. Herrmann, J. M., Koll, H., Cook, R. A., Neupert, W. & Stuart, R. A. (1995) *J. Biol. Chem.* **270,** 27079–27086.
33. Wischmann, C. & Schuster, W. (1995) *FEBS Lett.* **374,** 152–156.
34. Kadowaki, K., Kubo, N., Ozawa, K. & Hirai, A. (1996) *EMBO J.* **15,** 6652–6661.
35. Kubo, N., Harada, K., Hirai, A. & Kadowaki, K. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9207–9211.
36. Kubo, N., Jordana, X., Ozawa, K., Zanlungo, S., Harada, K., Saski, T. & Kadowaki, K. (2000) *Mol. Gen. Genet.* **263,** 733–739.
37. Figueroa, P., Gomez, I., Holuigue, L., Araya, A. & Jordana, X. (1999) *Plant J.* **18,** 601–609.
38. Figueroa, P., Gomez, I., Carmona, R., Holuigue, L., Araya, A. & Jordana, X. (1999) *Mol. Gen. Genet.* **262,** 139–144.
39. Adams, K. L., Daley, D. O., Qui, Y.-L., Whelan, J. & Palmer, J. D. (2000) *Nature (London)* **408,** 354–357.